

Лабораторна робота №7

ІІЗ-21-5 Богайчук Денис

Мета: набути навичок працювати з даними і опонувати роботу у Python з використанням теореми Байєса.

Завдання 1. Ретельно опрацювати теоретичні відомості:

- Теорема Байєса

Теорема Байєса допомагає обчислити ймовірність певної гіпотези з урахуванням наявних даних.

Формула:

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Де:

$P(H|E)$ — це апостеріорна ймовірність гіпотези ННН, враховуючи наявність даних ЕЕЕ.

$P(E|H)$ — це правдоподібність, тобто ймовірність отримання даних ЕЕЕ, якщо ННН є правдивою.

$P(H)$ — це апріорна ймовірність гіпотези ННН.

$P(E)$ — це ймовірність спостереження даних ЕЕЕ.

Ця теорема широко використовується у машинному навчанні, зокрема для задач класифікації та оцінки ризиків.

- Типи наївних байєсівських класифікаторів

Розрізняють кілька видів наївних байєсівських класифікаторів, які підходять для різних типів даних:

Гаусівський (Gaussian Naive Bayes):

Використовується для безперервних даних, припускаючи, що вони мають нормальний (гаусівський) розподіл.

Мультиноміальний (Multinomial Naive Bayes):

Застосовується для текстових даних, зокрема для класифікації текстів, наприклад, аналізу тональності. Використовується для дискретних змінних, таких як частота появи слова.

Бернуллівський (Bernoulli Naive Bayes):

Використовується для задач із бінарними даними, коли враховується лише наявність або відсутність певної ознаки.

- Де використовується Наївний Байєс

Метод активно використовується в різних галузях:

Текстова класифікація: Фільтрація спаму, аналіз емоційного забарвлення тексту, категоризація документів.

Медицина: Виявлення хвороб на основі симптомів.

Фінанси: Оцінювання платоспроможності та виявлення шахрайських операцій.

Рекомендаційні системи: Прогнозування уподобань користувачів.

Обробка зображень: Розпізнавання об'єктів та класифікація зображень.

Завдання 2. Ретельно розібрати приклад: прогнозування з використанням теореми Байєса.

Розглядається задача класифікації, де ймовірності розраховуються на основі умовних залежностей між характеристиками та класами.

Розрахунок для класу "Yes":

Обчислюються ймовірності для кожної умови:

$$P(\text{Outlook} = \text{Rain} \mid \text{Yes}) = \frac{3}{9}$$

$$P(\text{Humidity} = \text{High} \mid \text{Yes}) = \frac{3}{9}$$

$$P(\text{Wind} = \text{Weak} \mid \text{Yes}) = \frac{6}{9}$$

$$P(\text{Yes}) = \frac{9}{14}$$

Тоді:

$$P(\text{Yes}) = P(\text{Outlook} = \text{Rain} | \text{Yes}) \times P(\text{Humidity} = \text{High} | \text{Yes}) \times P(\text{Wind} = \text{Weak} | \text{Yes}) \times P(\text{Yes})$$

Розрахунок для класу "No":

Аналогічно розраховуються ймовірності для "No":

$$P(\text{Outlook} = \text{Rain} | \text{No}) = \frac{2}{5}$$

$$P(\text{Humidity} = \text{High} | \text{No}) = \frac{4}{5}$$

$$P(\text{Wind} = \text{Weak} | \text{No}) = \frac{2}{5}$$

$$P(\text{No}) = \frac{5}{14}$$

Тоді:

$$P(\text{No}) = P(\text{Outlook} = \text{Rain} | \text{No}) \times P(\text{Humidity} = \text{High} | \text{No}) \times P(\text{Wind} = \text{Weak} | \text{No}) \times P(\text{No})$$

Нормалізація результатів:

Після обчислення значень для "Yes" та "No", нормалізуємо їх:

$$P(\text{Yes} | \text{дані}) = \frac{P(\text{Yes})}{P(\text{Yes}) + P(\text{No})}$$

$$P(\text{No} | \text{дані}) = \frac{P(\text{No})}{P(\text{Yes}) + P(\text{No})}$$

Висновок: Теорема Байєса та наївний байєсівський класифікатор були досліджені як теоретично, так і практично, продемонструвавши їхню ефективність у розв'язанні задач класифікації через обчислення умовних і апостеріорних ймовірностей.